# Adopting spatial copula framework in regional frequency analysis

STAHY 2014
Statistical Hydrology Workshop

INRS
Université d'avant-garde

Masdar INSTITUTE


M. Durocher(1), F. Chebana(1) and T.B.M.J. Ouarda(2)

(1) Institut National de Recherche Scientifique, 490 rue de la Couronne, Québec, Canada (martin.durocher@ete.inrs.ca, fateh.chebana@ete.inrs.ca)

(2) Masdar Institue of Science and Technology P.O. Box 54224, Abu Dhabi, UAE (touarda@masdar.ac.ae)


IAHS AISH

## 1. Introduction

### Context

- **RFA is used for predicting flood quantiles at ungauged locations**. The rational is that two sites having similar hydrological properties should behave similarly. In that case, valuable information can be transfered between sites.
- **Hydrological dissimilarity** between a gauged site and an ungauged location cannot be calculated. Instead, a physiographical space in which an **associated metric** between site characteristics must be used.
- In **physiographical spaces**, flood quantiles can be predicted by **interpolation methods**, such kriging.

### Problematic

- Usually, flood quantiles share a **log-log relationship** with site characteristics, which **creates bias and suboptimal prediction with traditional kriging** techniques.
- Traditional kriging does not account for **non-constant variance**.
- **The problem** associated with traditional kriging technique **can be resolved by considering Spatial Copula**, an extension of traditional geostatistical framework where spatial dependance is characterized by a copula.

## 2. Case study

### Hydrological data

- Specific flood quantiles of 100-year return period

### At-site analysis

- Southern Quebec, Canada
- Number of sites :151
- Natural flow regime
- Record length: >15 years

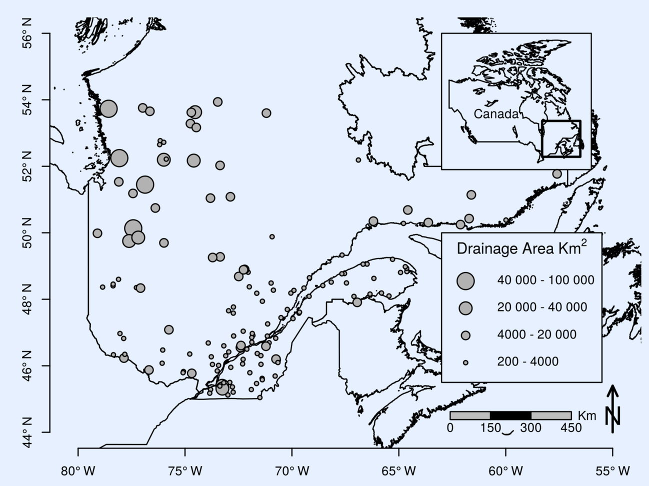

**Fig 1 : Map of the stations**

## 2. Spatial copula

- A multivarite distribution $G$ can be expressed as

$$G(\mathbf{x}) = C\left[(F_1(x_1), \ldots, F_n(x_n)\right]$$

where $\{F_i\}_{i=1}^{n}$ are margins for $\mathbf{x}' = (x_1, \ldots, x_n)$ and $C$ is a copula

- For spatial analysis, the copula has the same dimension as the number of site $n$ and **the strengh of the dependance** must be associated with a **distance** $h$.
- Spatial copula must allows for **strong dependance**

$$C_h \to M^n \quad \text{when} \quad h \to 0$$

and perfect **independence**

$$C_h \to \Pi^n \quad \text{when} \quad h \to \infty$$

- In spatial copula framework, the margins of the distribution $G$ are treated separately from its dependence. Hence, there is 2 set of parameters : the **marginal part** $\eta$ and the **copula part** $\theta$.
- Joint estimation of both part can be performed by optimizing a pairwise likelihood

$$L(\mathbf{z} \mid \eta, \theta) = \prod_{i<j} f(z_i, z_j \mid \eta, \theta)$$

where $\mathbf{z}' = (z_1, \ldots, z_n)$ are spatial observation and $f$ is the bivariate density of two sites $i$ and $j$.

- For known parameters $(\hat{\eta}, \hat{\theta})$, **the plug-in predictive distribution** (PPD) at ungauged location is the product of the marginal density and the conditional copula [2] :

$$p(z \mid \mathbf{z}, \theta) = f_{\hat{\eta}}(z) \times c_{\hat{\theta}}\left[F_{\hat{\eta}}^{-1}(z) \mid \mathbf{w}\right]$$

where $\mathbf{w}' = (w_1, \ldots, w_n)$ and $w_i = F_{\hat{\eta}}^{-1}(z_i)$

- Predictors can be calculated from the mean or the median of the PPD. For instance, the median is the quantity $F_{\hat{\eta}}^{-1}(w^*)$ for which

$$1/2 = \int_0^{w^*} c_{\hat{\theta}}(u \mid \mathbf{w})du$$

## 3. Physiographical space

- There is no general agreement on what hydrological similarity between sites should be. Here the **dissimilarity** between two sites $i$ and $j$ is defined as the **hydrological distance**

$$h_{i,j} = d(Z_i, Z_j)$$

between the vectors of flood quantiles $Z_i = (Z_{i,1}, \ldots, Z_{i,r})$ with return periods $1, \ldots, r$

- Let a physiographical space be a subspace of **coordinates**

$$S_i = AX_i$$

where $X_i$ are site characteristics and $A$ is a matrix that projects $X_i$ on the physiographical space.

- **A metric is associated with a dissimilarity measure** if

$$\text{Small } d(S_i, S_j) \to \text{Small } d(Z_i, Z_j)$$

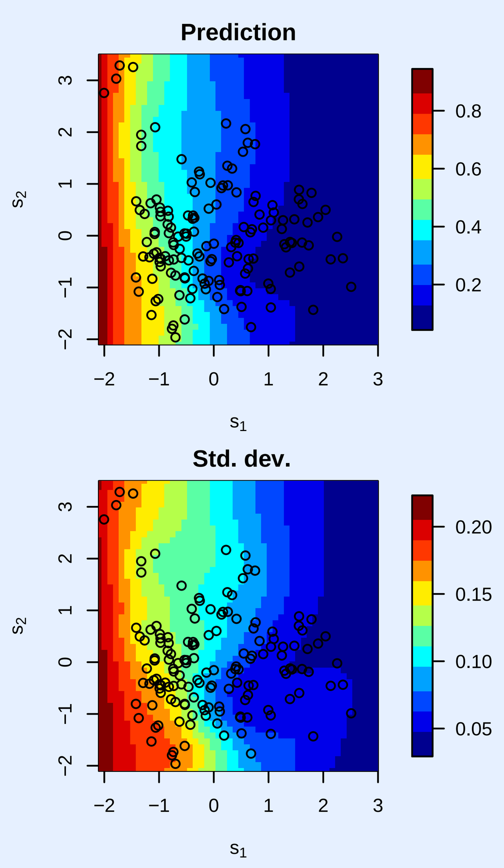


**Fig 2 : Predictions in physio. space**

- A way to build $A = (\mathbf{a}_1, , \mathbf{a}_m)$ is by using canonical correlation analysis, for which canonical pairs

$$\mathbf{s}_k = \mathbf{a}_k X \quad \text{and} \quad \mathbf{v}_k = \mathbf{b}_k Z$$

sequentially optimize $\mathrm{corr}(\mathbf{s}_k, \mathbf{v}_k)$ for $k = 1, \ldots, m$

## 4. Model

### Marginal part($\eta$)

- **Regional distribution** of the flood quantiles is **log-normal**.

$$log(Z_i) \to N\left[\mu(S_i), \sigma^2(S_i)\right]$$

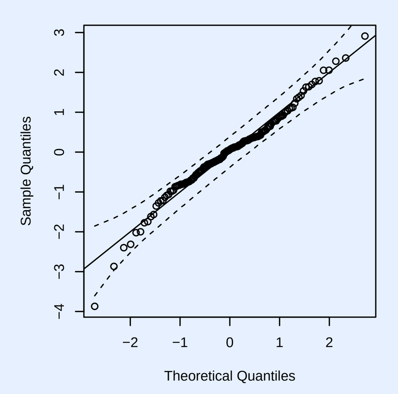


**Fig. 3 : Normalized QQ-plot**

- **A linear trend** is added to account for the strong correlation between the first canonical coordinates $S_{i,1}$ and the flood quantiles:

$$\mu(S_i) = \beta_{\mu,0} + \beta_{\mu,1}\ S_{i,1}$$
$$\sigma(S_i) = \beta_{\sigma,0} + \beta_{\sigma,1}\ S_{i,1}$$

### Copula part ($\theta$)

- The spatial dependance is characterized by a **Gaussian copula with pairwise correlation**

$$\rho(S_i, S_j \mid \lambda, \tau) = (1 - \tau)\ exp\left[-3\frac{d(S_i, S_j)}{\lambda}\right]$$

where $\lambda > 0$ (practical range) controls the correlation as $d(S_i, S_j) \to \infty$ and $\tau$ is a local measurement error (nugget effect)

- The **Goodness-of-fit** test [1] based on binned pairwise observations validates the choice of a Gaussian copula. For each bins, P-values are larger than 20% are found.

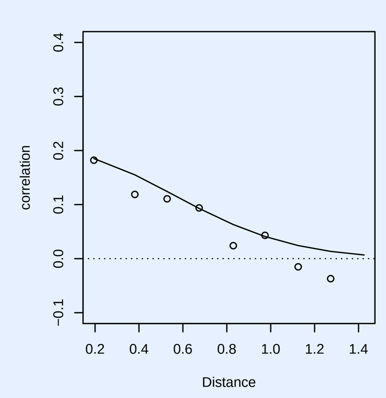


**Fig 4 : Correlogram QS100**

## 5. Results

- The performance of the prediction obtained by spatial copula is assessed by **Leave-one-out cross-validation**. In turn each gauged station is considered as ungauged and a predicted value is obtained as the **median of the PPD**.
- The analysis of the residuals shows the presence of large relative discrepancies (Fig. 5-Left). Note the presence of **problematic stations** previously identified for this database [3].
- The absolute residuals at logarithm scale (Fig. 5-Right) show a **decreasing variance** that is coherent with the trend $\sigma(S_i)$

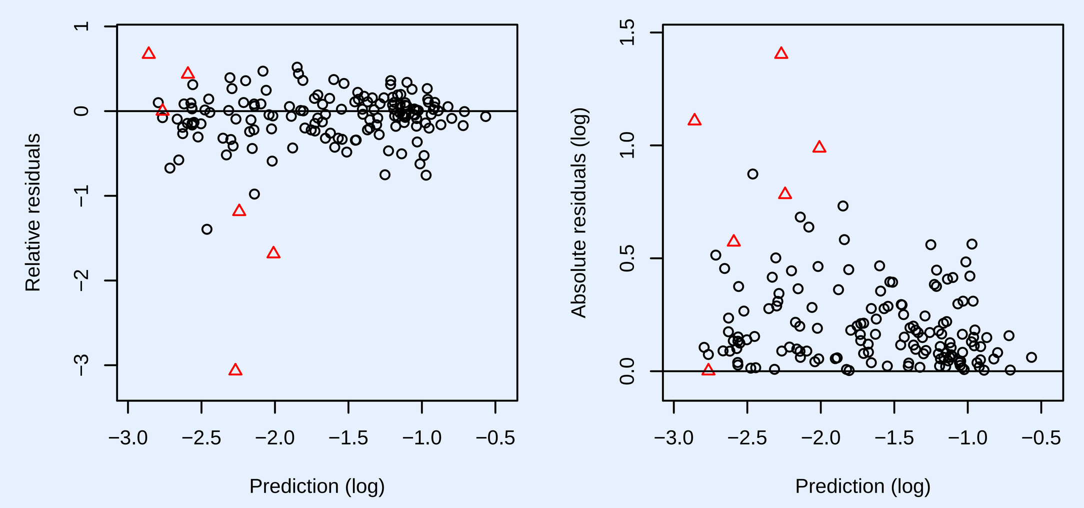


**Fig 5 : Residuals**

- **The prediction power** of the spatial copula approach (SCop) is compared to other methods on the same database:
  - Multiple regression with CCA-delineation (CCA)
  - Residual drift kriging (Krig)
  - Generalized additive model (GAM)
  - Artificial neural networks (ANN)

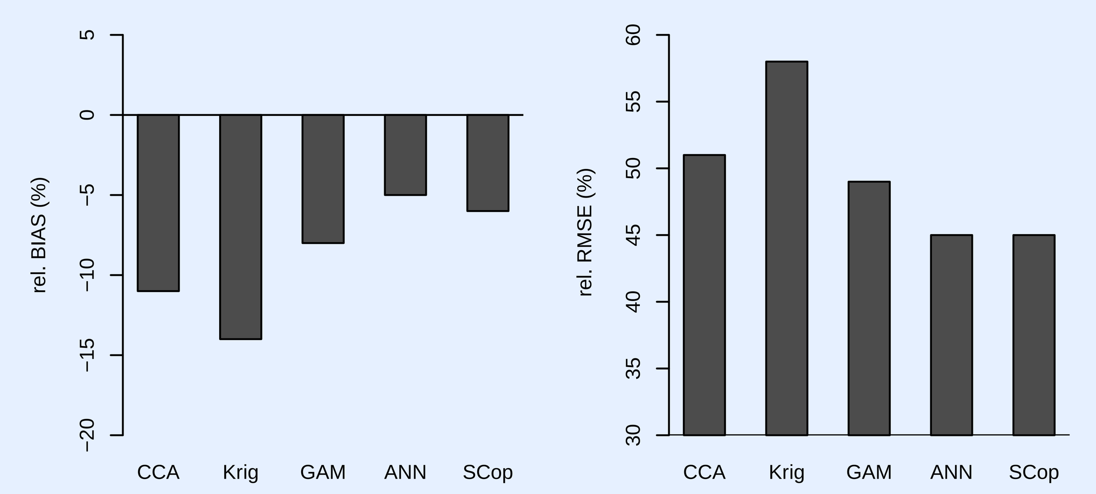


**Fig 6 : Performance criteria**

- In comparison with traditional kriging (Krig), the results of spatial copula (Scop) is associated with an important **reduction of the relative bias**.
- Overall, Scop and ANN have the **best** *rel.RMSE* from the methods considered here.

## 7. Conclusion

- The spatial copula framework offers a full probabilistic model that can **account for non-constant variance**.
- The spatial copula framework appears more **appropriate in presence of problematic stations** in comparison with traditional kriging.
- The spatial copula framework has **competitive performance with the best method**. In particular, it improves over traditional kriging.
- The important **relative bias** observed with the traditional kriging approach **is reduced** greatly with the spatial copula approach.

## References


[1] Bárdossy, A. (2006) Copula-Based Geostatistical Models for Groundwater Quality Parameters. Water ressour res, 42(11).

[2] Bárdossy, A., and Li, J. (2008) Geostatistical Interpolation Using Copulas. Water ressour res, 44(7).

[3] Chokmani K, Ouarda TBMJ (2004) Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. Water ressour res. 40(12).


## Acknowledgments


Financial support was provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada.